NASA CR-170568

(NASA-CR-173011) THE UARS AND OPEN DATA SYSTEM CONCEPT AND ANALYSIS STUDY. EXECUTIVE SUMMARY (Engineering and Economics Research, Inc.) 36 p HC A03/MF A01 CSCL 05B

N83-34847

Unclas

G3/82 36088

ENGINEERING & ECONOMICS RESEARCH, INC.

UARS AND OPEN DATA SYSTEM CONCEPT AND ANALYSIS STUDY

EXECUTIVE SUMMARY

Prepared for:

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

Under Contract Number: NAS 5-26962

Submitted by:

M. Mittal
J. Nebb
H. Woodward

Engineering and Economics Research, Inc.
1951 Kidwell Drive
Vienna, Virginia 22180

March, 1983

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

This report offers alternative concepts for a common design for the UARS and OPEN Central Data Handling Facility (CDHF) (see Section 4). The designs are consistent with requirements shared by UARS and OPEN (Section 2) and the data storage and data processing demands of these missions (Section 3). Because more detailed information is available for UARS, the design approach has been to size the system and to select components for a UARS CDHF, but in a manner that does not optimize the CDHF at the expense of OPEN. Costs for alternative implementations of the UARS designs are presented in Sections 4.1 and 4.2, showing that the system design does not restrict the implementation to a single manufacturer. Processing demands on the alternative UARS CDHF implementations are then discussed in Section 4.3. With this information at hand together with estimates for OPEN processing demands (Section 3.2), it is shown that any shortfall in system capability for OPEN support can be remedied by either component upgrades or array processing attachments rather than a system redesign.

In addition to a common system design, it is shown in Section 5 that there is significant potential for common software design, especially in the areas of data management software and non-user-unique production software.

The report then gives cost examples for several modes of communications between the CDHF and Remote User Facilities (Section 6). The report concludes with a discussion of the potential application of technologies expected to reach fruition before the mission timeframe (Section 7).

2.0 UARS AND OPEN SYSTEM LEVEL REQUIREMENTS

Based upon available documentation and input from GSFC technical personnel, a list of OPEN and UARS missions system level requirements, assumptions and intercomparisons was generated for the report, in which particular emphasis was placed upon the Central Data Handling Facility (CDHF). Based upon this information, it is seen that there are a number of system level functions common to both a UARS and an OPEN CDHF. The major of these common functions are:

- o Data ingest of playback data
- o Routine production processing of the data
- o Data management
- o Investigator communications.

The distribution of these functions within the proposed CDHF concepts is defined in Section 4.

2.1 UARS and OPEN Systems Elements

Not only do the UARS and OPEN CDHF's share similar system level requirements, but their relations to institutional facilities are also similar. This is illustrated in Figure 2.1-1. As shown in the figure, in addition to the CDHF, both UARS and OPEN ground systems consist of the following functional elements:

- Data capture
- Orbit determination
- Attitude determination
- Command management
- Payload operations control

ORIGINAL PAGE IS
OF POOR QUALITY

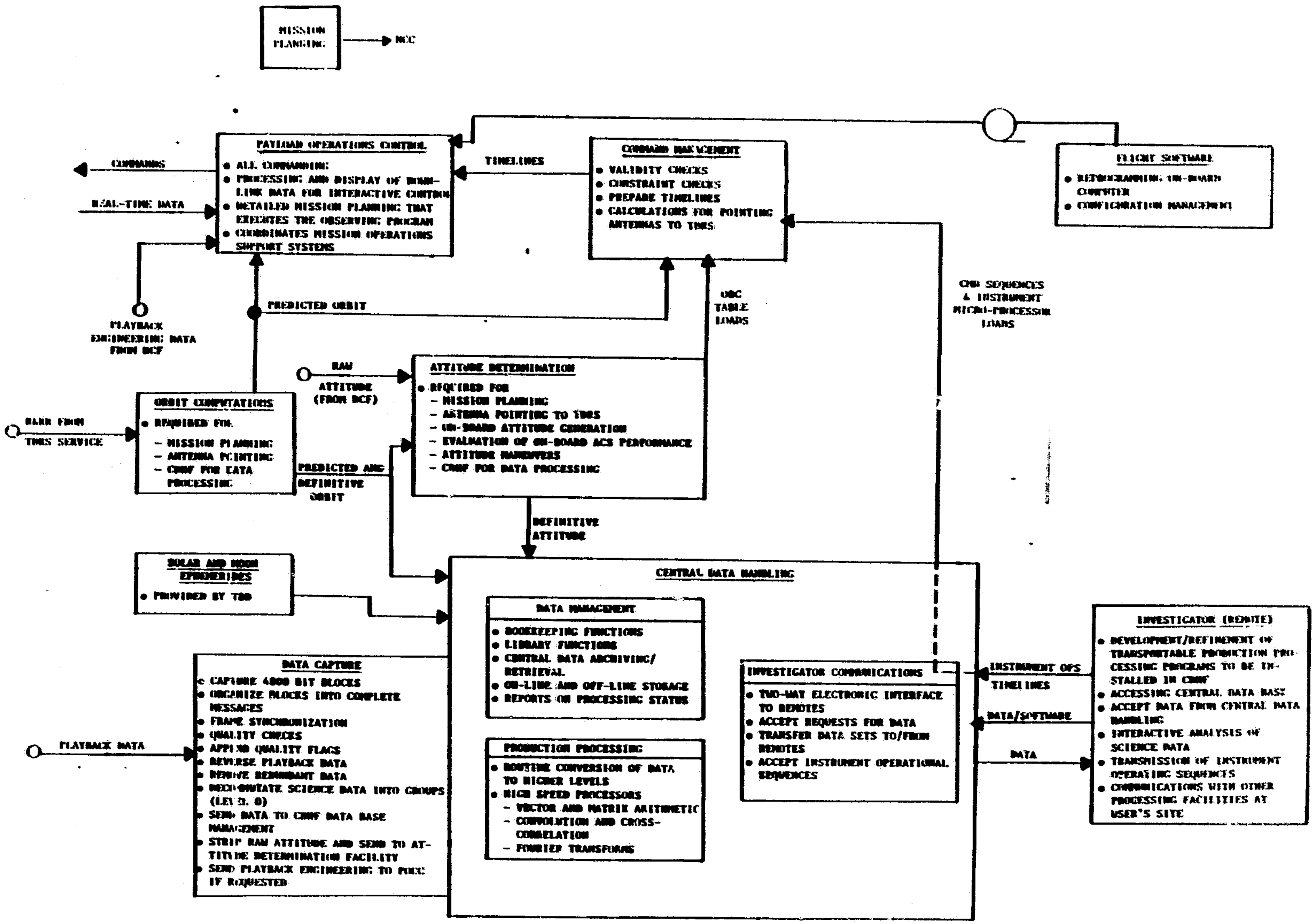


FIGURE 2.1-1 OPEN-UARS GROUND SYSTEM ELEMENTS

- Flight software
- Mission planning
- Communications

Additionally, the PI's will be provided interactive remote facilities suitable for analysis of the processed data.

The main functions performed by the ground system elements as well as the inter-relationships among them are shown in Figure 2.1-1.

3.0 DATA STORAGE AND PROCESSING ANALYSIS

In order to derive system concepts for processing and managing data within the CDHF, estimates for their data storage and data processing requirements are necessary. These are presented in Sections 3.1 and 3.2, respectively.

3.1 UARS and OPEN Storage Analysis

Available information allows for an analysis of the requirements for data processing and storage. It should be noted, however, that the information available for UARS is more complete.

Table 3.1-1 presents UARS data volumes for UARS production processing by data category. It is seen that daily production data volume is about 740 Mbytes. An on-line storage capacity of about 111 Gbytes of production data as well as an additional 9 Gbytes of support data and PI data submissions from remote sites will be required.

Table 3.1-2 presents the OPEN storage requirements. These requirements have been derived from information presented in the proposals for the OPEN instruments which have been selected. It is seen that daily production data volume is about 1.2 Gbytes and that an on-line storage capacity of about 418 Gbytes will be required.

3.2 UARS and OPEN Processing Estimates

In order to derive system concepts for the CDHF, not only must the data requirements be at hand but it is also necessary to focus upon the magnitude of the processing demands upon the CDHF. These are presented for UARS and OPEN in the following paragraphs.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 3.1-1 UARS INSTRUMENT DATA STORAGE REQUIREMENTS

INSTRUMENT	AVG. Data Rate (Kbps)	Daily Volume (MB) Level				Daily Total	ON-LINE STORAGE [1] (MB)
		0	1	2	3		
Winds and Temperatures (WINTERS)	1.3	14	14.5	14.2	2.8	45.5	9,755
High Resolution Doppler Imager (HRDI)	4.5	48.4	86	40	0.6	175.0	24,908
Cryogenic Limb Array Etalon Spectrometer (CLAES)	1.1	12	32	10.9	0.18	45.08	1,663
Halogen Occulation Experiment (HALOE)	1.1	15.8	12.5	2.7	0.04	31.04	2,013
Improved Stratospheric and Mesospheric Sounder (ISAMS)	0.5	5.4	2.7	0.8	8.3	17.2	5,049
Microwave Limb Scanner (MLS)	4.0	58.8	90	61	31	240.8	52,968
Particle Environment Monitor (PEM)	2.7	28.5	109.5	11 [2]	5 [2]	154.0	12,321
Solar Ultraviolet Spectral Irradiance Monitor (SUSIM)	1.0	10.7	0.5	0.23	0.01	11.44	251
Solar Stellar Irradiance Compar. Exper. (SOLSTICE)	0.1	0.7	0.5	0.5	0.02	1.72	303
Solar Backscatter Ultra-violet Radiometer (SBUV) [3]	0.32	2.4	2	2	0.4	6.8	1,380
(MAGNETOMETER) [4]	0.3	3.25	6.5	-	-	9.75	116
TOTAL FROM ALL INSTRUMENTS		199.95	356.7	133.33	48.35	738.33	110,807

[1] 10 Day L0; 30 Days L1; 540 Days L2 and L3
[2] 10:1 decrease in data volume from L1 → L2 and 2:1 decrease from L2 → L3 (estimated from PI requirement for graphics data)
[3] Similar to instrument flown on advanced TIROS-N Series.
[4] Supplied by PEM and used by PEM experiment only.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 3.1-2 OPEN DATA STORAGE REQUIREMENTS

Lab	PRINCIPAL INVESTIGATOR (PI) Name	Institution/ Organization	ID	Experiment	Instrument Name	Down-Link Data Rate Kbps	Duty Cycle %	Average Data Rate Kbps	Daily Data Volume MB	DAILY DATA VOLUME (MB) L0	L1[4]	L2[5]	Total	On-Line Storage MB[6]	Off-Line Storage Requirement MB[7]	Need for Array Processor
PPL	Russell	UCLA	81	MAG. FIELDS		0.5	100	0.5	5.4	5.4	16.2	3.24	24.84	1,944	18,133.2	
PPL	Mozer	UCB	85	ELEC. FIELDS		2.5 5.0	90 10	2.75	29.7	29.7	89.1	17.82	136.62	10,692	99,732.6	
PPL	Shawhan	U of Iowa	25	PLASMA WAVE INSTRUMENT		[3]		14.64	179.7	179.7	539.1	107.82	826.62	64,692	608,432.6	
PPL	Scudder	GSFC	26	HOT PLASMA	HYDRA	4.4	100[2]	4.4	47.5	47.5	142.5	28.5	218.5	17,100	159,508.0	
PPL	Shelley	LPARL	24	HOT PLASMA COMPOSITION		4.35	100[2]	4.35	47.0	47.0	141.0	28.2	216.2	16,920	157,836.0	
PPL	Chappell[1]	MSFC	31	COLD PLASMA	TIDE	2.5	100[2]	2.5	27.0	27.0	81.0	16.2	124.2	9,720	90,666.0	
PPL	Higbie[1]	LASL	43	ELEC. PARTICLE	CEPPAD	3.2 9.6	90 10	3.84	41.47	41.47	124.4	24.88	190.75	14,928	139,247.5	
PPL	Fritz[1]	NOAA/SEL	53	ELEC. PARTICLE COMPOSITION	CAMMICE	1.152	100	1.152	12.45	12.45	37.35	7.47	57.27	4,482	41,807.1	
PPL	Feldman	JHU	90	VISIBLE IMAGER	MAID	16.2	80	12.96	140.0	140.0	420.0	84.0	644.0	50,400	470,120.0	YES
PPL	Torr	Utah	92	UV IMAGER	AIME	TBD	TBD	20	216.0	216.0	648.0	129.6	993.6	77,760	725,328.0	YES
PPL	Imhof	LPARL	64	X-RAY IMAGER	PIXIE	3.0	80	2.4	25.92	25.92	77.76	15.55	119.23	9,331	87,039.9	
PPL							LAB TOTAL	71.49	772.1	772.1	2,316.3	463.3	3,551.66	277,956	2,592,711.8	
EML	McPherron	UCLA	66	MAG. FIELDS		0.5	100	0.5	5.4	5.4	16.2	3.24	24.84	1,944	27,199.8	
EML	Maynard	GSFC	59	ELEC. FIELDS		1.4 2.0	90 10	1.46	15.8	15.8	47.4	9.48	72.68	5,688	79,586.6	
EML	McIlwain	UCSD	74	HOT PLASMA	EFIELDS	2.0	100	2.0	21.6	21.6	64.8	12.96	99.36	7,776	108,799.2	
EML	Scarf	TRW	91	PLASMA WAVE INSTRUMENT		2.138 6.158	90 10	2.54	27.44	27.44	82.32	16.46	126.22	9,878	138,210.9	YES
EML	Parks	U of Wash.	19	HOT PLASMA		8.2	100[2]	8.2	88.56	88.56	265.68	53.14	407.38	31,882	446,089.1	
EML	Burch	SRI	45	HOT PLASMA COMPOSITION		2.0 5.0	10 1	0.25	2.7	2.7	8.1	1.62	12.42	972	13,599.9	
EML	Chappell[1]	MSFC	31	COLD PLASMA	TIDE	2.5	100[2]	2.5	27.0	27.0	81.0	16.2	124.2	9,720	138,999.0	
EML	Higbie[1]	LASL	43	ELEC. PARTICLE	CEPPAD	3.2 9.6	90 10	3.84	41.47	41.47	124.4	24.88	190.75	14,928	208,871.25	
EML	Fritz[1]	NOAA/SEL	53	ELEC. PARTICLE COMPOSITION	CAMMICE	1.466	100	1.466	15.8	15.8	47.4	9.48	72.68	5,688	79,586.6	
EML							LAB TOTAL	22.75	245.7	245.7	737.1	147.42	1,130.22	88,452	1,237,590.9	
GTL	Lepping	GSFC	33	MAG. FIELDS		0.46 3.7	90 10	0.995	10.31	10.31	30.93	6.18	47.42	3,711	51,986.7	
GTL	Mozer[1]	UCB	84	ELEC. FIELDS		0.6 1.2	90 10	0.66	7.1	7.1	21.3	4.26	32.66	2,556	35,762.7	
GTL	Gurnett	U of Iowa	44	PLASMA WAVE INSTRUMENT		1.974 10.0	100 5	1.674	18.08	18.08	54.24	10.85	83.17	6,509	91,071.15	YES
GTL	Frank	U of Iowa	77	HOT PLASMA	LEPEDEA	3.5	100	3.5	37.8	37.8	113.4	22.68	173.88	13,608	190,398.6	
GTL	Williams	NOAA/SEL	63	ELEC. PARTICLE COMPOSITION	EPIC	2.7	100	2.7	29.16	29.16	87.48	17.49	134.13	10,497	146,872.35	
GTL							LAB TOTAL	9.49	102.6	102.6	307.2	61.44	471.06	36,864	515,788.8	
IPL	Behannon	GSFC	36	MAG. FIELDS		0.46 3.7	90 10	0.955	10.31	10.31	30.93	6.18	47.42	3,711	51,924.9	
IPL	Kaiser	GSFC	35	PLASMA WAVE INSTRUMENT		0.515 6.96	90 10	1.16	12.52	12.52	37.56	7.51	57.59	4,507	63,061.05	
IPL	Ogilvie	GSFC	50	HOT PLASMA		0.40	100[2]	0.40	4.32	4.32	12.96	2.59	19.87	1,555	21,757.65	YES
IPL	Gloeckler	U of MD	86	HOT PLASMA COMPOSITION		0.471	100[2]	0.471	5.08	5.08	15.24	3.05	23.37	1,829	25,590.15	
IPL	Lin	UCB	13	ELEC. PARTICLE		0.380	100[2]	0.380	4.10	4.10	12.3	2.46	18.86	1,476	20,651.7	
IPL	McDonald	GSFC	54	COSMIC RAYS	EPACT	0.25	100	0.25	2.7	2.7	8.1	1.62	12.42	972	13,599.9	
IPL	Teegarden	GSFC	28	GAMMA RAYS		0.193	100	0.193	2.08	2.08	6.24	1.25	9.57	749	10,479.15	
IPL							LAB TOTAL	3.81	41.1	41.1	123.3	24.66	189.06	14,796	207,020.7	
							TOTALS FOR ALL FOUR LABS	107.54	1,161.3	1,161.3	3,483.9	696.78	5,341.98	418,068.0	4,553,112.2	

[1] Repeats
[2] Assumed
[3] 1.5 Kbps @ 100% duty cycle; 35.2 Kbps @ 100%; and 128 Kbps @ 5% (256 Kbps for 30 minutes; or 25.6 KHz for 4 hours)
[4] Increase of 3:1 from L0 → L1 (assumed)
[5] Decrease of 5:1 from L1 → L2 (assumed)
[6] 100 Days L1 + 100 Days L2
[7] 36 months each for EML, GTL, IPL; 24 months of PPL

Based upon the UARS data processing requirements contained in the actual questionnaire responses submitted by the PIs and the CSC study which summarizes and synthesizes these responses it is estimated that in order to process a day's production data for the instruments selected a total processing load of about 48,000 Million Floating Point Operations (MFLOPS) would be required. If these operations could be spread uniformly over an 8-hour period (one shift) then the effective throughput of the computing machinery would be 1.68 MFLOPS/sec. In other words, CPU sizing for processing should be in the 2 MFLOPS/sec (effective throughput) range. This estimate does not include I/O and data management demands.

Information for OPEN data processing which is comparable to the results in the CSC study has not yet been developed. However, gross estimates can be made for OPEN by extrapolating what is known about UARS together with analyzing the selected OPEN instrument proposals. When this is done, it is estimated that the ratio of OPEN processing demands to UARS processing demands is about 5.2:1. Thus CPU sizing for OPEN data processing is about in the 9 MFLOPS/sec range (effective throughput), excluding I/O and data management demands. The analysis for deriving this estimate is presented in the report.

4.0 CDHF SYSTEM CONCEPTS

This section presents two system design approaches for satisfying the requirements of either a UARS or an OPEN CDHF. Because the UARS CDHF is assumed to precede the OPEN CDHF, the overall approach has been to size a system and select components for a UARS CDHF, but in a manner that does not optimize the CDHF for UARS at the expense of OPEN. Indeed. the shortfall in system capability for OPEN support could be remedied by component upgrades rather than a system redesign.

In what follows, a detailed analysis is made for UARS. System upgrades to accommodate OPEN are indicated in Section 4.4.

Based upon available information, the following major UARS functions have been identified:

- o Data Ingest and L-0 Production
- o L-0 to L-1 Production
- o L-2 to L-3 Production
- o Data Services To/From Remotes
- o Remote Batch
- o Data Management

Based upon these functions, two functional concepts for a UARS CDHF have been formulated. The first concept presented is a CDHF featuring dual mainframe systems. The second concept presented is a CDHF configuration featuring a single mainframe system. Neither concept depends upon unique hardware subsystems available from only a single vendor. The dual mainframe and single mainframe concepts are described in Sections 4.1 and 4.2, respectively, and two different hardware implementations of each concept

are presented. Summary information regarding significant features and costs are presented in Tables 4.0-1 and 4.0-2.

4.1 Dual Mainframe Concept

In the dual mainframe concept the various CDHF functions are carried out by two autonomous software compatible mainframes which share a common data base, and the CDHF functional workload is split between a Production Processor (PP) system and a Data Manager/Processor (DM/P) system as illustrated in Figure 4.1-1. As indicated in Figure 4.1-1, the extensive arithmetic and matrix manipulation services required to accomplish daily L-2 production and to provide remote batch services are provided by the PP and its associated array processing facilities, while the computationally less demanding L-0, L-1 and L-3 production services, as well as the (primarily) non-arithmetic data ingestion, data management and remote site interface services are provided by the DM/P.

The PP and DM/P would be sized to permit the processing of a day's volume of UARS data in one work shift, with capacity to spare. The PP would be sized in the 3 to 3.5 MIPS range, while the less powerful DM/P would operate in the range of 1 MIPS.

Since the dual mainframe concept features two independent software compatible mainframes sharing a common database, certain backup capabilities are inherent in this approach which are not present in a single mainframe approach. In the event of PP outage, the DM/P and array processing facilities may be used to carry on UARS production at a reduced rate of approximately 50% (2 work shifts, with little or no margin). In the event of DM/P outage, the PP can assume the responsibilities of the DM/P and complete all daily processing tasks within 2 work shifts.

TABLE 4.0-1

ALTERNATE IMPLEMENTATION FEATURE SUMMARY

DUAL MAINFRAME IMPLEMENTATION		SINGLE MAINFRAME IMPLEMENTATION	
IBM[1]	CDC[2]	IBM[3]	CDC[4]
* Production Processor • IBM 3033-N-8 - 8 Mbyte Memory - 3 MIPS	* Production Processor • CDC Cyber 170 Series 700 170-730 Dual Processor - 262K x 60 Bit Memory - 3.5 MIPS		
• 2 FPS AP-109L Array Processors - 6 MFLOPS Average (ea) - 256K Memory (ea)	• CDC Advanced Flexible Processor (For Array Processing) - 200 Million Arithmetic Operations/Second (Avg)	* System Processor • IBM 3081 - 16 Mbyte Memory - 10.4 MIPS	* System Processor • CDC Cyber 170 Series 700 170-760 Processor - 262K x 60 Bit Memory - 11 MIPS • Extended Memory - 1 Million 60 Bit Words
* Data Manager/Production Processor • IBM 4341-L02 - 8 Mbyte Memory - 0.75 MIPS	* Data Manager/Production Processor • CDC Cyber 170 Series 700 170-720 Processor - 262K x 60 Bit Memory - 1.2 MIPS * Shared Extended Memory • 1 Million 60 Bit Words		
* On-Line Mass Storage[5] (Shared, not disk) • 2 IBM 3851-A04 - 472 Gbytes Total	* On-Line Mass Storage[6] (Shared, not disk) • 2 MASSTOR 860 - 440 Gbytes Total	* On-Line Mass Storage[5] (not disk) • 2 IBM 3851-A04 - 472 Gbytes Total	* On-Line Mass Storage[6] (not disk) • 2 MASSTOR M860 - 440 Gbytes Total

__Notes__:

[1] Partial listing; complete listing in Table 4.1-3.

[2] Partial listing; complete listing in Table 4.1-6.

[3] Partial listing; complete listing in Table 4.2-3.

[4] Partial listing; complete listing in Table 4.2-5.

[5] System limit; may not be expanded.

[6] Expandable.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4.0-2

SYSTEM COST SUMMARY ESTIMATES

FACILITY	DUAL MAINFRAME IMPLEMENTATION		SINGLE MAINFRAME IMPLEMENTATION	
	IBM[1]	CDC[2]	IBM[3]	CDC[4]
Central Data Handling Facility (CDHF)	$8,463,662	$7,405,105	$9,539,135	$8,236,388
Software Compatible Remote Sites[5] (1 site/19 sites)	$208,890/ $3,968,910	$747,837/ $14,208,903	$208,890/ $3,968,910	$747,837/ $14,208,903
Combined Cost of CDHF and 19 Remote Sites	$12,432,572	$21,614,008	$13,508,045	$22,445,291

Notes:

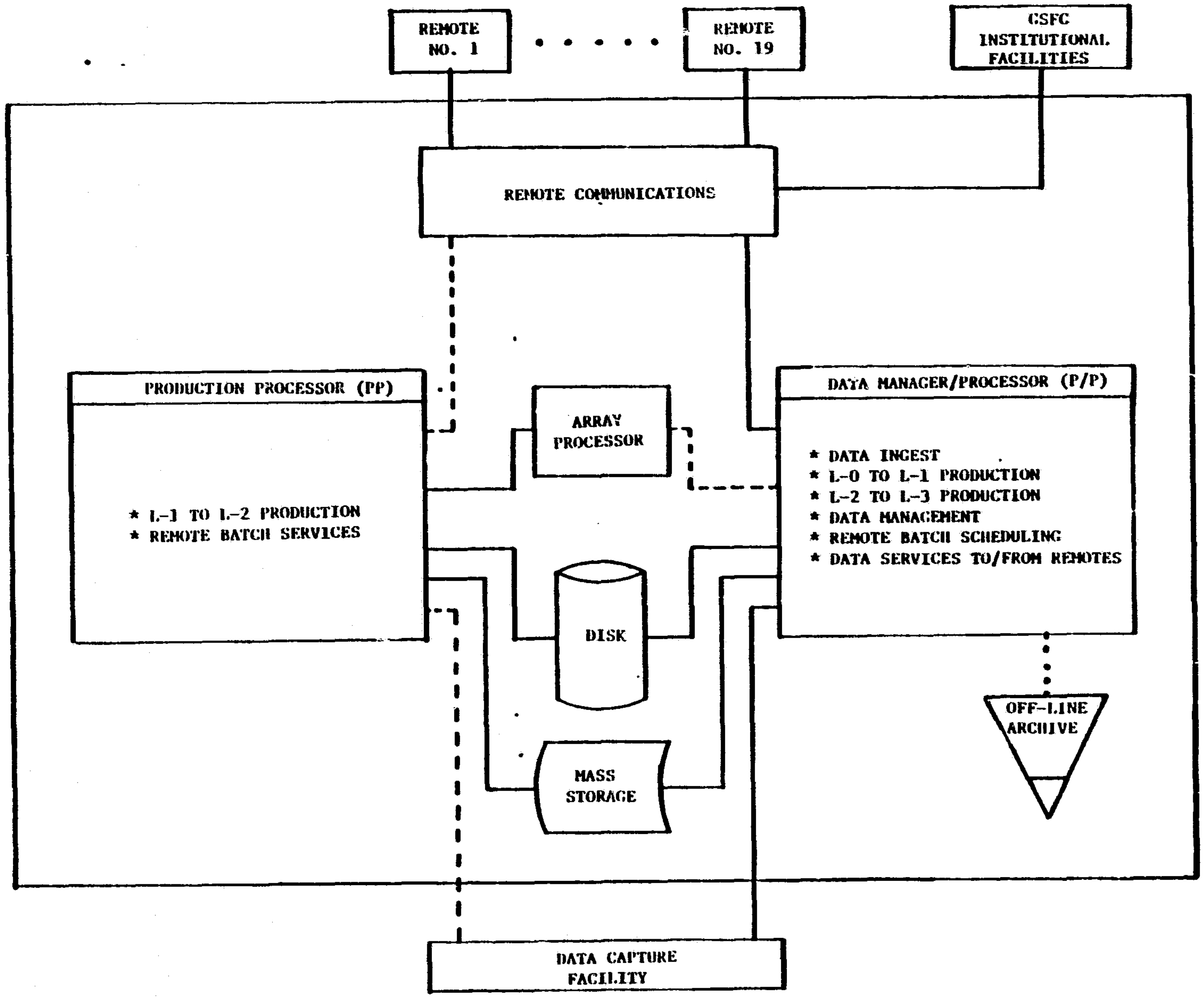
[1] Detailed cost estimates presented in Tables 4.1-3 and 4.1-4.

[2] Detailed cost estimates presented in Tables 4.1-6 and 4.1-7.

[3] Detailed cost estimates presented in Tables 4.2-3 and 4.1-4.

[4] Detailed cost estimates presented in Tables 4.2-5 and 4.1-7.

[5] CDC remote facilities have extensive computational capabilities appropriate for OPEN. IBM remote facilities, while less powerful, are more appropriate for UARS. The CDC remote facilities represent the low end of the software compatible Cyber 170 Series 700 equipment line.



LEGEND:
------ DENOTES BACKUP (REDUNDANT) PATH

FIGURE 4.1-1 DUAL MAINFRAME FUNCTIONAL CONCEPT

Two possible hardware implementations of the dual mainframe concept have been prepared. The first implementation features IBM mainframes, while the second implementation features CDC mainframes. Both implementations require array processors. Tables 4.0-1 and 4.0-2 summarize these two implementations and their costs.

4.2 Single Mainframe Concept

The single mainframe concept accomplishes all of the CDHF functions using a single large mainframe. This concept is illustrated in Figure 4.2-1.

As was the case with the dual mainframe concept, the single mainfrme is sized to permit the processing of a day's volume of UARS data in one work shift. However, no array processing is required. In contrast to the dual mainframe concept, however, the single mainframe concept does not include the capability to operate at reduced levels in the event of mainframe failure since there is no mainframe redundancy.

Two hardware implementations of the single mainframe concept were derived: IBM and CDC. Tables 4.0-1 and 4.0-2 summarize these implementations and their costs.

4.3 Production Processing Demands/Estimates for UARS

Table 4.3-1 summarizes the minimal input, output and processing resources that would be consumed by the dual mainframe and single mainframe implementations. The values listed in this table are minimal since operating system resource demands and system inefficiencies are not included.

4.4 OPEN/UARS CDHF Commonality

Since the on-line storage required for OPEN is about 418 Gbytes (see

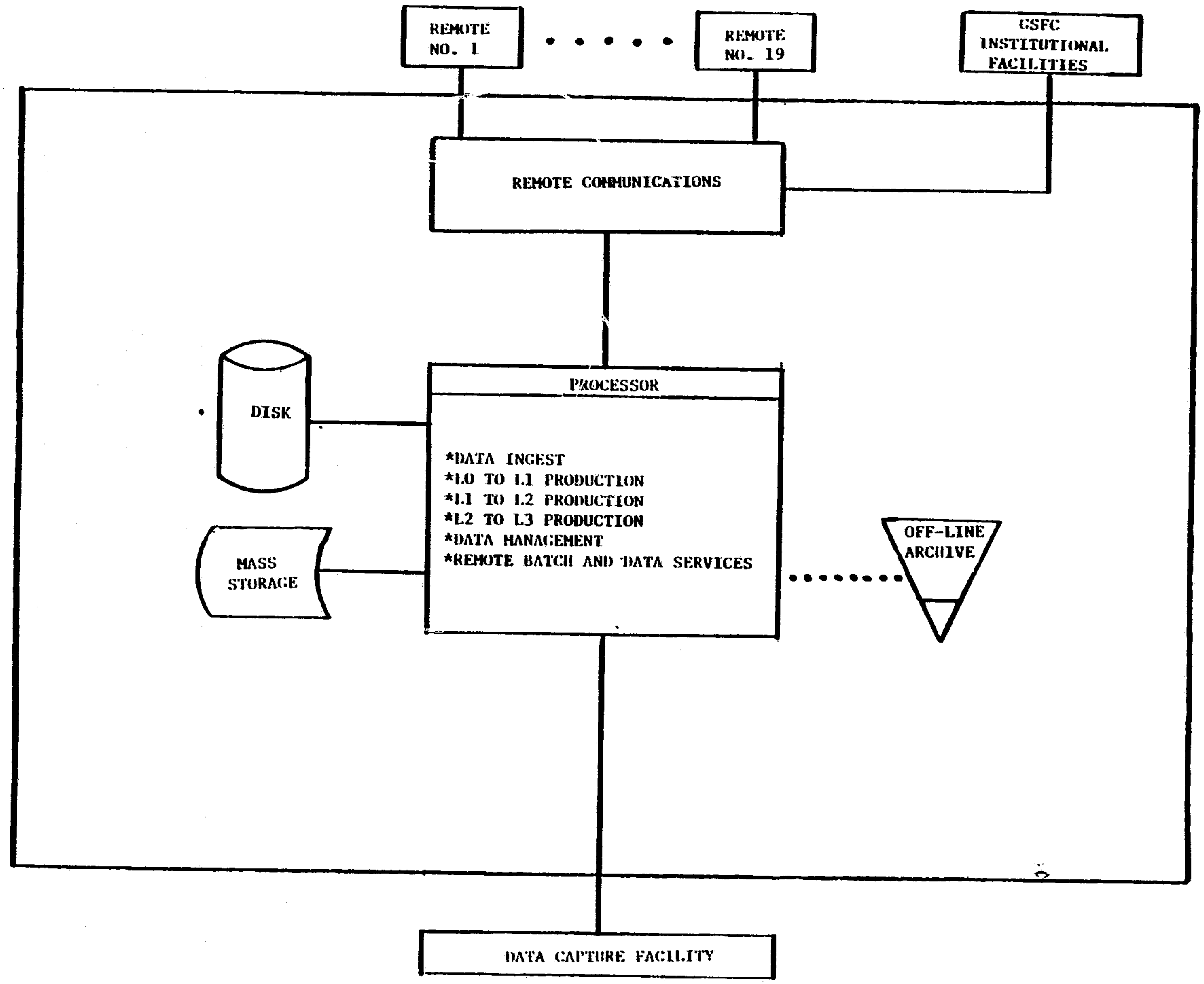


FIGURE 4.2-1 SINGLE MAINFRAME FUNCTIONAL CONCEPT

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4.3-1

MINIMUM PRODUCTIVE RESOURCE DEMANDS (HOURS)

CATEGORY		IBM DUAL MAINFRAME		CDC DUAL MAINFRAME		IBM SINGLE MAINFRAME	CDC SINGLE MAINFRAME
		DM/P	PP	DM/P	PP		
INGEST:	READ RAW	0.79	–	0.79	–	0.79	0.79
	PROCESS	0.59	–	0.37	–	0.04	0.04
	WRITE L-0	0.56	–	0.56	–	0.56	0.56
L-1 PROC:	READ L-0	0.56	–	0.56	–	0.56	0.56
	PROCESS	2.25	–	1.41	–	0.16	0.15
	WRITE L-1	1.00	–	1.00	–	1.00	1.00
L-2 PROC:	READ L-1	–	1.00	–	1.00	1.00	1.00
	PROCESS	–	2.38[1]	–	2.04[2]	2.56	2.42
	WRITE L-2	–	0.37	–	0.37	0.37	0.37
L-3 PROC:	READ L-2	0.37	–	0.37	–	0.37	0.37
	PROCESS	1.10	–	0.69	–	0.08	0.08
	WRITE L-3	0.13	–	0.13	–	0.13	0.13
TOTALS:	INPUT[3]	1.72	1.00	1.72	1.00	2.72	2.72
	PROCESS	3.94	2.38	2.47	2.04	2.84	2.69
	OUTPUT[3]	1.69	0.37	1.69	0.37	2.06	2.06

Note:

[1] In addition, concurrent array processing consumes 3.16 hours of AP190L(1) resources and 3.31 hours of AP190L(2) resources.

[2] In addition, concurrent array processing consumes 0.78 hours of Advanced Flexible Processor (AFP) resources.

[3] "Ingest; READ RAW" estimate based on 1.4619×10^9 bits input at 512K bits/sec; all other "READ/WRITE" estimates use an estimate of 10 usec per byte.

Table 3.1-2), this is in the range of the mass storage systems envisioned for the UARS CDHF. Thus, major UARS system upgrades are only required to accommodate the higher OPEN processing load. The upgrade could be accomplished as follows: for the dual mainframe approach, the Production Processor (PP) would be substantially upgraded; for the large single mainframe approach, array processors would be added. The latter approach appears to be the more straightforward and appears to offer the greater potential for achieving of hardware and software commonality. An explanation is given in the paragraphs the follow.

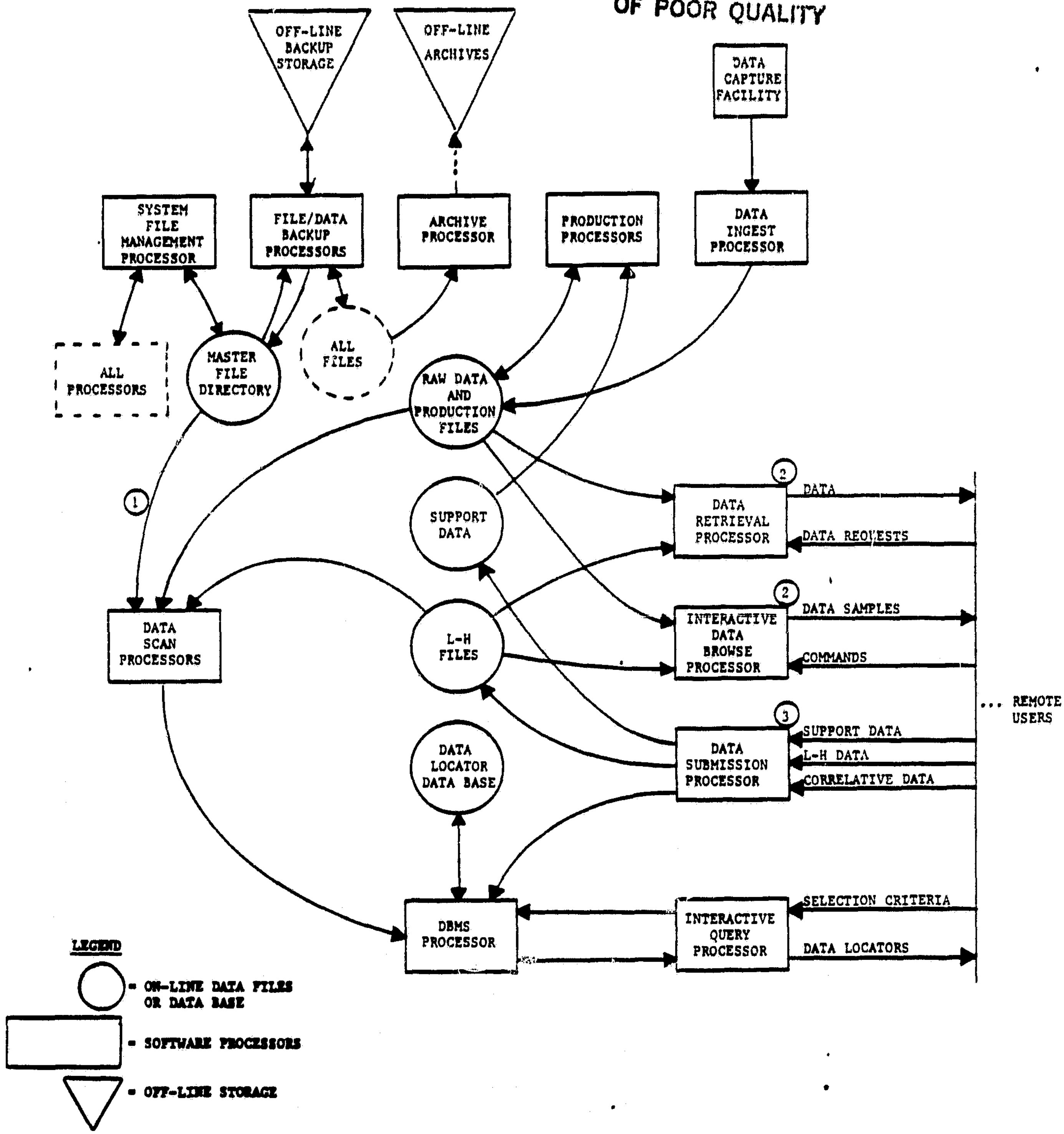
Recall that for UARS, it is felt that a computer in the 10-11 megainstructions/sec range could accommodate all UARS processing, with no attached array processor required, in about 3 hours (theoretical throughput). Since the OPEN processing load is estimated to be about 5.2 times that of UARS (Section 3.2), it would appear that about 15.6 hours of the UARS mainframe would be required for OPEN. However, the attachment of array processors to the UARS single mainframe offers promise for significantly reducing the 15.2 hours demand on the computer. If this is the case, without substantial hardware design, commonality could be achieved.

In addition to the common hardware design inherent in this approach, there could be promise for achieving a measure of software commonality. As will be seen in Section 5, there appear to be substantial areas of commonality between the OPEN and UARS software systems both in the areas of data management software and the production software. If both OPEN and UARS processing utilized the same mainframe, then the software would be available to both and substantial cost savings could be realized.

5.0 DATA PROCESSING AND MANAGEMENT CONCEPT

This section presents a unified data processing and management concept for OPEN and UARS. The approach taken is cognizant of the challenges offered by managing the massive volumes of data stored on-line at the respective CDHF's. Among these challenges are access speed, data base recovery and data base reorganization. Furthermore the concept presented here satisfies investigator browse and retrieval requirements as well as the need to manage massive volumes of sequential files. In the discussion potential areas for using commercially available products are indicated.

A description of a common approach to UARS and OPEN data management is summarized pictorially in Figure 5.0-1. The concept presented incorporates standard system sequential file processors (vendor utilities) to manage the large quantities of UARS or OPEN data at the CDHF and would thus minimize storage overhead for these data. Use of sequential files would also simplify any data restoration process required to recover data destroyed as the result of system malfunctions. Since large sequential files do not lend themselves to rapid querying by remote users, a Data Locator Data Base (DLDB) designed for fast access would be provided to assist users in locating data of interest. The DLDB would be event and condition oriented and would be of a coarser time granularity than the UARS and OPEN data that is summarized. Such a data base, being a summary of the data elements used to derive it, could be much more compact and rapidly accessible than would be a data base which consists of the constituent data elements of the events themselves. Vectors into specific files that contained data corresponding to particular events or conditions would be filed in the DLDB



NOTES:
1. FILE ATTRIBUTES (TIME OF CREATION, LAST MODIFICATION, ETC.)
2. "READ ONLY" ACCESS.
3. ANY USER MAY SUBMIT DATA, A USER MAY ONLY UPDATE/DELETE HIS OWN DATA.

FIGURE 5.0-1

CDHF DATA PROCESSING AND MANAGEMENT CONCEPT

to provide the user with the information necessary to access, browse through or retrieve data of interest.

5.1 Production Cycle

Similarities between UARS and OPEN suggest that a common software production framework and collection of production software utilities could be configured for use at both CDHF's.

As pictured in Figure 5.0-1, a typical production cycle would begin with the arrival of spacecraft data via the data capture facility. Raw data would be read into the CDHF data processing system by the Data Ingest Processor, subjected to elementary quality control checks, and stored. Subsequently the Production Processor (PP) would be activated in turn. These production processors need not be unique to a particular CDHF. During the production processing the various PP's would input some level of data together with any required spacecraft or instrument oriented support data and produce specific higher level outputs. As various segments of the production process are completed, appropriate Data Scan Processors (DSCAN) would be activated. The function of the DSCAN processors would be to examine the various new production files registered in the system Master File Directory and to develop (predefined) summary information for incorporation into the Data Locator Data Base (DLDB). As an example, the DSCAN might note at which point in time a peak reading occurred in a particular subsystem and record such items as the value of the reading; the (real) time and date at which the event occurred, instrument status; the name of the data file containing the reading; and the relative position (within the data file) of the reading. Subsequently the DSCAN would submit this (and other) significant events data to the Data Base Management System

Processor for incorporation into the DLDB. Once incorporated into the DLDB, the significant events and vectors into the production data file system would be available to the user community via the Interactive Query Processor.

5.2 User Interface

Four processors are provided to allow users to submit, locate and retrieve data (see Figure 5.0-1). These processors are as follows: a Data Submission Processor for permitting users to submit higher level data and support and correlative information; an Interactive Query Processor for determining availability and location of data; an Interactive Browse Processor for examining selected data fields; and a Data Retrieval Processor for forwarding specific data to the user site.

5.3 Data Security

Since the data stored at the CDHF will represent significant expenditures of manpower, analytic and data processing resources, it is essential that the CDHF include the necessary elements of security to protect data from accidential or deliberate loss or destruction. The concept presented in Figure 5.0-1 includes provisions for data security that minimizes the chances for the loss or destruction of data by providing off-line backup copies of data and by controlling accesses to on-line data.

The creation of the off-line backup copies of on-line data would be provided using File/Data Backup Processors. While processors could be especially written for the CDHF, in many cases off-the-shelf system utilities are available to provide backup file copies on magnetic tape. The creation of off-line backup data copies (probably using magnetic tape

as a backup medium) could be an ongoing process throughout the lifetime of the CDHF. Furthermore, in the event of system software or hardware failures or user errors resulting in the loss of on-line data, an appropriate File/Data Backup Processor could re-establish the data on-line from the most recent backup copy available for that data. Such a recovery method would avoid the necessity of lengthy (multi-hour, in some cases) computer runs to recreate lost data from lower levels data.

Since the CDHF system will be a multi-user system, access to on-line data will of necessity be limited to a certain degree (at a minimum it would be necessary to prohibit concurrent updating of the same data by multiple users and/or processors). In terms of Figure 5.0-1, these types of access control could be provided by the System File Management and DBMS processors. User identification codes or account numbers augmented (if necessary) by privilege passwords would probably prove adequate.

5.4 Archive Function

Depending upon the nature and compatibility of the archive medium and/or interface, the Archive Processor shown in Figure 5.0-1 would prepare copies of archive data, using a medium such as computer compatible tape. While archive medium generation for production data could be postponed until the end of the CDHF lifetime, consideration should be given to an ongoing archive process (perhaps a daily or weekly archive generation run) throughout the CDHF lifetime. An ongoing archive process (of data which have become static) could preclude the necessity for an extended series of archive production runs involving the transfer of hundreds of billions of bytes of data from on-line storage. The introduction of the concept of an ongoing archive process might also lead to significant savings in time and

resources if it proved feasible to combine certain elements of the archive process with an ongoing data backup process.

6.0 COMMUNICATIONS COSTS; CDHF/REMOTES

In order to derive cost estimates for CDHF/Remote communications a comparison was made for the following three modes:

- o Packets
- o Digital Service Leased Line
- o Satellite Hop

The comparison was made under the following assumptions:

- o Remote located 2000 miles from GSFC
- o 12 Mbytes/day of traffic (average) between GSFC and a UARS Remote
- o 38.5 Mbytes/day of traffic (average) between GSFC and an OPEN Remote.

Under these assumptions a table was derived which presents a summary of the communications costs to the remotes (see Table 6.0-1). The cost figures for Packets and Digital Service Leased Line were derived from Fundamentals of Data Communications by Jerry Fitzgerald and Tom. S. Eason, 1978, as well as conversations with cognizant GSFC personnel. The satellite communications costs were based on Planning Research Corporation (PRC) System Services Company's NASCOM Circuit Regression, which appears in Development of NASA DMS Performance/Cost Models, dated 5 January 1982.

TABLE 6.0-1

COMMUNICATIONS COSTS (DOLLARS/MONTH/REMOTE)

Communication Mode	Costs (Dollars/Month/Remote)	
	UARS	OPEN
Packets	$18,768	$59,274
Leased Line	$2,139	$2,139
Satellite (Domestic)	$3,370	$3,370
Satellite (Overseas)	$19,430	$19,430

7.0 POTENTIAL TECHNOLOGY APPLICATIONS

The technology for implementing the UARS and OPEN data systems exists at the present time. However, there are technologies that should be available during the mission time-frames that could be utilized for a more cost effective or better performing system. The technologies examined are in the areas of data management, mass storage, software language development and communications.

7.1 Data Management

The most promising potentially applicable data management technology is that of the data base machine. Until recently data base management systems (DBMS's) have been software systems which executed on standard general purpose computers. However, two major limitations have surfaced under this implementation scheme. Data management systems that run on conventional computers run into bottlenecks when processing a large volume of transactions on very large (10 Gbytes) data bases. This is due to the data staging "bottleneck" between mass storage and main memory. The second limitation is that users are continually demanding more sophisticated DBMS capabilities such as backup and recovery, integrity and security controls, etc. These capabilities are needed by OPEN and UARS and require tremendous overhead. Consequently, a number of researchers have proposed the use of dedicated or specialized processors to execute data management functions. These are called data base machines (DM).

Several DM architectures are under investigation. All involve parallelism in one form or another and therefore take advantage of emerging VLSI technology. An example of a DM available today is a Britton-Lee computer designed specifically for DBMS processing. With software and

hardware the entire system can be purchased for about $200,000 (cost will vary depending on data base size and options). It is capable of data base access times equivalent to those obtained on a 5 to 10 MIPS standard computer with a software data base and there are plans to increase performance by another 5 to 10 fold. It will currently support up to 10 Gbytes of disk storage.

7.2 Mass Storage of Data

A common theme to the OPEN and UARS architectures discussed in this volume is that to achieve a balance between operational performance and system costs a hierarchy of computer memories/storage technologies is required. This hierarchy consists of a spectrum of cache/main memory, mass storage, and archival memory devices that span roughly six orders of magintude in both performance and cost. Because most technology involved in the existing memory hierarchy continues to reduce the per-bit storage cost at about the same rate, there will be no cross-over within the hierarchy within the near future. Therefore, memory hierarchies will continue to play a key role in the design of cost effective system architectures.

The storage technologies for accomplishing the objectives of the OPEN and UARS missions are well at hand. However, although there are numerous choices which can be made among alternate computer systems for performing production and communication tasks, there are only two choices for implementing the mass storage function. These choices, describe previously in this report are the IBM Mass Storage System (MSS) and the Masstor Virtual Storage System (VSS). Both these systems are basically automated magnetic tape-cartridge read/write systems that access the appropriate

cartridge, load it, and transfer the data to a staging disk in a matter of seconds. In the near term it does not appear that these devices will be supplanted. However, it can be anticipated that with the continuing price decrease in VLSI technology, more device intelligence will be built into mass storage devices. This would help remove the data-location burden from the CPU as well as minimize I/O traffic between mass storage and main memory. Additionally there could be an implementation in the mass-storage devices of such features as format initialization, limit checking, data compression and expansion, and error correction.

On the horizon the only apparent alternative to the magnetic cartridge mass storage devices seems to be the emerging optical disk storage systems. Optical disks promise a higher storage density and a lower per-bit cost than any other mass storage medium. Additionally they are they are made of materials that can be stored for many years without stringent environmental controls. However, opitcal disks suffer the drawback of being write-once devices. Although most magnetic tape is used in a write-once manner, there is a reluctance to utilize a new technology that forces this mode of operations.

At the present time, RCA has completed experimental opitcal disk systems that can record 5 Gbytes of data on one side of an optical disk at rates exceeding 100 Mbits/sec. These systems have provided a bit error rate of one-in-100 Mbits and can access any block of data in less than 0.5 seconds. There are plans to design a unit that would hold a number of optical disk platters that would be retireved and loaded as the need arose. It is planned that the worst case access time for a data block in this system would be about 5 seconds to retrieve data from a stored disk and .5

seconds if the disk were already on line. This type of system has been proposed to have 1.25 terabytes of storage.

Before optical data storage hardware becomes a reality, however, much work remains in the mechanics, the optics and the recording medium. Nonetheless, the current level of development activities suggests that operational systems will become widespread by the late 1980's or early 1990's.

A mass storage system that is exclusively optical disk does not appear feasible for OPEN and UARS-type projects because the write-once limitation could lead to a database size of over a terabyte. However, reversible data (Levels 0 and 1 for UARS and Level 0 for OPEN), which would rarely be altered, could be optically stored. An example of an advantage here could be the ease by which large quantities of this data could be recorded on a single disk (5 Gbytes or more) and sent by an express package service to the investigators. This could relieve a heavy I/O and communications burden from the CDHF.

7.3 Software Language Developments

The most likely major transition in languages that can be expected in the near future is the acceptance and use of the Ada language. Ada is currently under development by the Department of Defense (DOD) to be used in all of their software systems. Not only is it a powerful and flexible structured language, but it also serves as a program support environment, particularly for transportability, as well as supplying a methodology for life cycle software development, particularly in the area of configuration management. The use of Ada for OPEN and UARS would require massive

programmer retraining, the positive features of Ada may not outweigh this initial disadvantage. Moreover the cost benefit of Ada has not yet been proved.

Currently there are research efforts under way for producing compilers for automatic program generation in the sense that languages would be produced which would allow statements about what the program is to do to generate high-level language algorithms for stating how the program is to produce the desired results. For languages under current research, the compiler determines the sequence of procedures by analyzing the statements entered. This is in contrast to conventional languages in which control flow is built into the program itself.

At the present time there are no commercially available compilers for automatic program generation. It does not appear that one would be available for OPEN and UARS. Moreover, it remains to be seen whether increased hardware performance can overcome the potential slowness and inefficiency of multi-level compilers which first translate specifications into high-level languages and then into machine-language instructions.

7.4 Communications

Communications will be paced by advances in satellites and optical fibers.

In satellite communications, research in the areas of space diversity and time-division techniques, developments in antenna technology, sophisticated high-speed on-board switching, exploiting higher frequency sections of the spectrum, and on-board error detection and correction would provide for much broader wideband capabilities in space. However, under the communications traffic assumed by UARS, for example, recurring

satellite and terresterial communications costs were about the same. As long as rate structures are determined as a function of distance, this can be expected to remain the case. It remains to be seen if this policy will change.

Over the next few years local networks will be wire-based. If fiber is to compete, interfaces must be developed for fiber-optic systems that are compatible with coaxial networks such as Ethernet. Also, research is needed to define network topologies that best utilize fiber optics. Standards are now being established for defining a general class of terminal device for an optical fiber system. A goal would be the interchangability of the terminal device with a terminal device for a wire-based network.

Outside of local network applications, high-speed fiber optic buses may fill the need for fast parallel transfers between a mainframe and high-speed peripherals. This may serve to relieve any potential data staging bottlenecks between mass storage and main memory, as could be the case in the CDHF.